Grenze International Journal of Engineering and Technology, January Issue

Hybrid Techniques for Object Detection & Tracking: A Review

Bhawna Narwala[1], Anil Gargb[2] and Shikha Bhardwajc[3]
[1]Research Scholar, Department of Electronics and Communication Engineering, Maharishi Markandeshwar Deemed to be University, Mullana, Haryana, 133207, India
Email: bhawnanarwal9@gmail.com
[2]Associate Professor, Department of Electronics and Communication Engineering, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar Deemed to be University, Mullana, Haryana, 133207, India
[3]Assistant Professor, Department of Electronics and Communication Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, 136119, India

***Abstract*— In this paper, a detailed analysis about a hybrid object detection and tracking framework that combine traditional image processing methods with modern deep learning models has been discussed. Techniques such as Kalman filtering, optical flow, and deep networks like YOLO and Faster R-CNN are combined to improve performance in real-time, multi-object scenarios. Some key challenges including occlusion, identity preservation, and environmental variability have been discussed in this paper. Some datasets like Pascal VOC, ImageNet, Microsoft COCO, etc. and parametric evaluations like precision, accuracy, mAP and MOTA etc. have been studied.**

***Keywords*— Computer vision, Image Processing, Object Recognition, Object Tracking.**

I. INTRODUCTION

The increasing apprehension regarding security and safety is inspiring an explosion of research focused on an auto-mated video analysis for object detection and tracking, resulting in a rapid advancement in hardware facilities such as camera processing machines and mobile phones. A key area of study in computer vision research is object identification & tracking, which aims to identify and track objects across sequences of frames and describes their behaviour [1] [2]. Currently, most everyday environment's, including streets, shopping centres, educational institutions, stations, and residences, are now under surveillance by a variety of technologically advanced monitoring systems. Detecting and tracking objects are essential processes in computer vision that allow machines to understand and analyze visual information from their environment, as illustrated in Fig. 1 [3].

The procedure of finding and recognizing things of interest within a video or photo frame is commonly referred to as recognizing an object. It returns bounding boxes surrounding objects along with their class labels such as "dog", "car", "bus". The importance of object detection is that to make machines understand a scene by detecting multiple objects and acts as the preliminary stage for more intricate systems, including tracking, behaviour analysis, or auto-mated decision-making [4]. Object detection techniques can be broadly categorized into two main groups: conventional and vast learning-based object exposure [5]. Deep learning methods can further sub-divide into two categories that are two-stage and one-stage detectors.

Grenze ID: 01.GIJET.12.1.198

Traditional object detection methods include Haar cascades, Deform-able Part-based Models (DPM), combination of Histogram of Oriented Gradients (HOG) & SVM. R-CNN, Mask R-CNN, Faster R-CNN etc are examples of two-stage detectors. SSD, YOLO etc are examples of one-stage detectors [6] [7]. The categorization of object detection techniques is represented in Fig. 2.

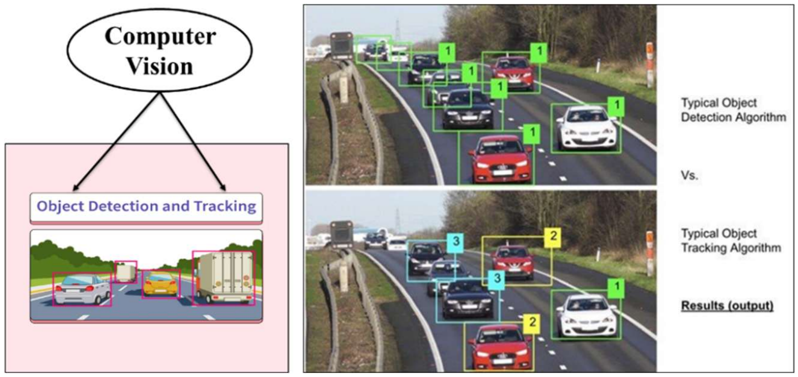


Fig. 1: Differentiate between Object Detection & Object Tracking [3]

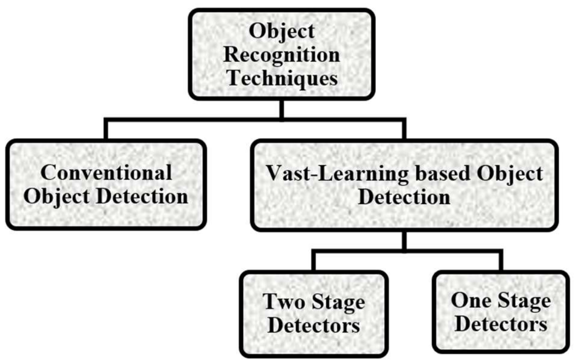


Fig. 2: Categorization of Object Recognition Techniques [7] [8]

Object tracking is a process that involves monitoring one or more detected objects throughout a series of frames in a video. It enhances detection by preserving the identity of objects over time [9]. The importance of object tracking is that they are essential for grasping temporal dynamics and to promote contextual and consistency understanding in video analysis [10] [11]. Object tracking tracks either single object tracking that is to monitor a single object as it moves through a video or multiple objects tracking that is to tracks numerous objects, addressing challenges such as re-identification and occlusion [12] [13]. Object tracking can be broadly categorized into three methods that are point based method, kernel tracking method and silhouette tracking method as shown Fig. 3.

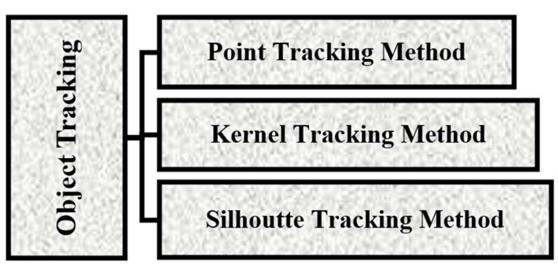


Fig. 3: Categorization of Object Tracking Techniques [14]

Fig. 4 illustrates the fundamental workflow of object finding and movement system. This diagram outlines the core process of object detection and tracking system. It starts with input data in the form of images or videos, which un-dergo pre-processing to improve clarity and uniformity [16] [17]. The enhanced data is then used for object detec-tion, where the system locates and classifies objects within each frame. Once detected, the tracking component fol-lows these objects across successive frames, maintaining their identities [18]. Lastly, performance evaluation is con-ducted to measure the system's effectiveness using metrics like precision, recall, and speed. These tasks are very crucial for real-world domains such as sports analytics, autonomous driving, robotics, surveillance systems, healthcare, augmented reality, and activity monitoring [19].

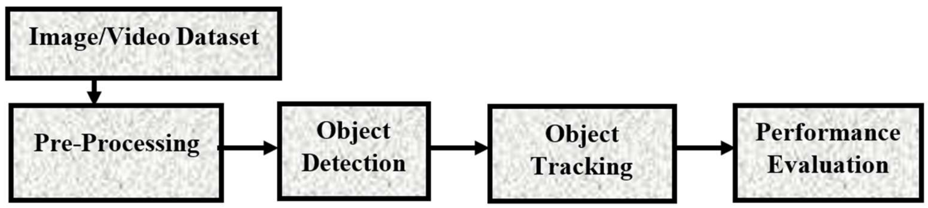


Fig. 4: Fundamental workflow of an Object Recognition and Tracking System [15]

The structure of the rest of this paper is organized as follows: Section 2 presents the significance and motivation behind object detection and tracking systems. Some of the object recognition as well as tracking techniques that were investigated in the earlier work are covered in Section 3. Section 4 introduces some datasets and evaluation metrics of object exposure and tracking. Finally, Section 5 discusses the conclusion and future trends.

II. MOTIVATION

Recognizing as well as tracking moving objects is a core task in computer vision, serving as the foundation for vari-ous real-life applications including activity monitoring, autonomous vehicles, and intelligent robotics [20]. But it involves various difficulties that can greatly influence the effectiveness and accuracy of tracking systems like occlu-sions, motion blur, varying lighting conditions, and the requirement to handle multiple objects at the same time [21] [22]. A major difficulty is occlusion; when an object is wholly blocked from view by other scene elements or partial-ly hidden from view, there can be a loss of tracking or misidentification especially in dense or complex scenes. An-other important challenge is achieving real-time processing because the algorithms need to process data quickly and efficiently (especially for high-resolution or high-frame-rate inputs) [23] [24] [25]. Data association continues to be a persistent issue, particularly when there are multiple moving objects involved; it becomes even more challenging when distinguishing between several similarly appearing objects or maintaining object identity across frames in crowded or rapidly changing scenes. Lastly, generalization remains problematic because models trained on particular datasets do not always generalize well in new or varied environments (e.g., different lighting conditions, meteorolog-ical conditions, object sizes, background) [26]. Tracking systems need to be able to adapt to a diverse range of real-world scenarios in order to be beneficial for everyday use [27]. The intention of this research endeavour is to ex-plore creative solutions, especially hybrid models made up from various algorithms, to contemporary problems in the rapidly emerging field.

III. LITERATURE SURVEY

N. H. Abdulghafoor et al. [28] (2022) accomplishes this by proposing a strong innovative algorithm to identify and follow objects from real-time camera-captured natural scenes. In this work an algorithm for tracking and detection that is sensitive to real and fundamental changes will be developed. The detection of numerous moving creatures, scarce resources and various difficulties are characteristics of this algorithm. This algorithm leverages the benefits of both principal component analysis (PCA) and vast learning networks to create an intelligent real-time tracking and recognition system. In order to improve performance over the recent detection and tracking algorithms it is done adaptively between the two approaches. Comparing the innovative algorithm to other tracking and recognition sys-tems and accomplishing high detection and classification accurateness demonstrated its efficacy and efficiency.

An enhanced system for identifying and monitoring aerial vehicles in inclement weather including snow fog and sandstorms is presented in A. S. Talaat et al. [29] (2023). The Retinex algorithm is used to improve images, the Yolov5 vast learning model is used to detect vehicles and Optical Flow, Kalman Filter & Euclidean Distance are used to track vehicles in order to increase overall accuracy. Corrected Optical Flow achieved the lowest tracking error (less than 5 pixels) and the results show evaluations of Precision and Recall along with a mean Average Precision (mAP) of 92.2%.

M. H. Ashraf et al. [30] (2023) proposes a vision-based VVM (Vehicle Velocity Monitoring) system that encom-passes tracking, velocity calculation and vehicle identification. The VVM system operates on a three-layer architec-ture and makes use of a solo RSU camera. The present CNN-based (Convolutional Neural Network) Fusion Vehicle Recognition Network is developed in the first layer to detect vehicles utilizing multiple stages and multi-scale fea-tures to handle changes in vehicle size and improve detection correctness. The hybrid vehicle recognition network and SORT (Simple Online Real Tracker) are used to monitor vehicles in the second layer. This tracking system makes use of the Hungarian algorithm for multi-vehicle composition and the Kalman filter algorithm to estimate the location of the vehicle. A technique for determining vehicle velocity on the moving axis is introduced in the third layer to address rapidity overestimation. Three datasets—PASCAL VOC for vehicle finding, UA-DETRAC for track-ing correctness and Urban Roadways for rapidity estimation—are used to assess the suggested methodology. The system attains 87.242% correctness in rapidity estimation, 88.72% mAP, & 84.35% MOTA.

Y. P. Huang et al. [31] (2023) demonstrated a novel fusion deep neural network framework that handles missing locations for pose estimation by combining BlazePose with an interpolation technique. Furthermore, it utilizes a net-work of long-short-term memory (LSTM), among for human motion recognition, a refined YOLOv4 algorithm for object exposure, or a Landmark Data Processing technique for the extraction of features. Correctly determining hu-man activity while ensuring optimal object utilization and movements is the overall objective of this integrated sys-tem. The model suggested obtained a processing speed of 10.47 frames per second, an action identification accuracy of 95.91% and mean average precision (mAP) of 97.68%.

H. M. Chuang et al. [32] (2024) present a multi-camera perception (MCP) and high-frequency perception (HFP) sys-tem for object recognition, which uses fusion algorithm to boost both exactness and speed, where the results show that MCP gives high exactness while HFP provides quick update rates.

The AIA-IFRCNN model was presented by K. Vijiyakumar et al. [33] (2024). It combines the Inception v2 model for feature extraction the DCF-CSRT algorithm for image annotation the Faster R-CNN for object detection and a softmax layer for image analysis. Three distinct data sets—Bird (Dataset 1), UCSDped2 (Dataset 2), & Under Water (Dataset 3)—were used to test the model's performance in terms of prediction accuracy, labeling time, center point proximity to the correct location and overlap between detected and actual objects. According to the results the AIA-IFRCNN model outperformed the other models in terms of both object detection accuracy and tracking velocity. Specifically, it outperformed other models with the highest accuracy of 95.62 percent on the Bird dataset. Addition-ally, it had the lowest average center location blunder with the Bird, UCSDped2 and Under Water datasets having respective averages of 4.16, 5.78, and 3.54. On those similar datasets it also had high overlay rates of 0.92, 0.90, & 0.94.

With emphasis on digital pathology analysis of images, M. Salvi et al. [34] (2020) reviewed several methods used in deep learning frameworks for improving input data via the preliminary processing and streamlining network out-comes through post-processing. Despite the review's focus on computerized pathology, lots of the methods of post- processing covered are also broadly relevant for various other areas of visual analysis.

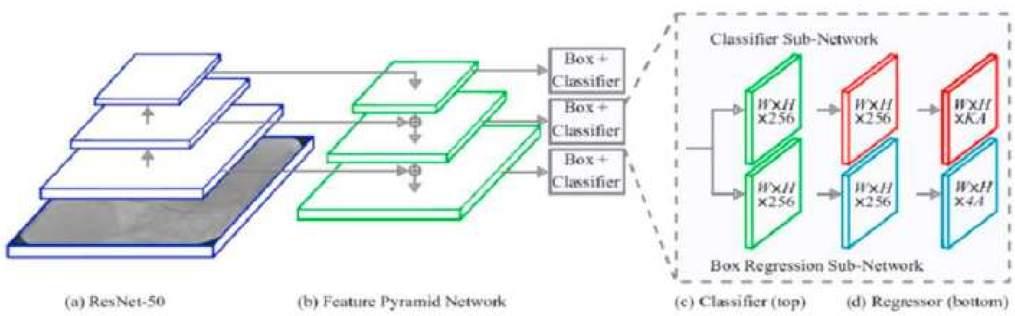
L. Liu et al. [35] (2023) suggested employing features taken from a deep convolutional neural network (DCNN) to track targets in dynamic scenes. They created a window-that-moves segmentation technique to assess how standardi-zation of data and dataset augmentation affected tracking performance. To investigate how data augmentation and normalization affect tracking performance, they developed an adaptive window segmentation approach. In order to boost feature selection, DCNN characteristics were subjected to Principal Component Analysis (PCA), and character-istics gathered from different network levels were compared. The research also investigated feature encoding tech-niques comparing locality-constrained linear coding to the Fisher Vector algorithm for encoding DCNN features. DCNN-based features outperform conventional feature fusion methods in terms of accuracy according to investiga-tional results. Additionally, the analysis demonstrates that the DCNN tracking algorithm remains more precise in a variety of illumination scenarios and performs well when partial obstruction is present.

J. Ding et al. [36] (2023) addressed frequent issues in dynamic environments by introducing a reliable characteristic threshold-based strategy to determining and segmenting moving objects, improved by better optical flow estimation. Unlike conventional optical flow and Otsu-based separation for fixed cameras, this technique employs a background feature separation with threshold strategy that integrates the LK (Lucas–Kanade) & HS (Horn–Schunck) optical flow techniques. An image pyramid paired with HS and LK techniques yields a high-precision, interference-resistant opti-cal flow approximation model. To handle motion occlusion, Delaunay triangulation is applied. The proposed seg-mentation method is then used on the optical flow field, leveraging Harris features and an image background affine transformation model to accurately extract moving objects. The last front region of the moving target is generated using morphological processing of images.

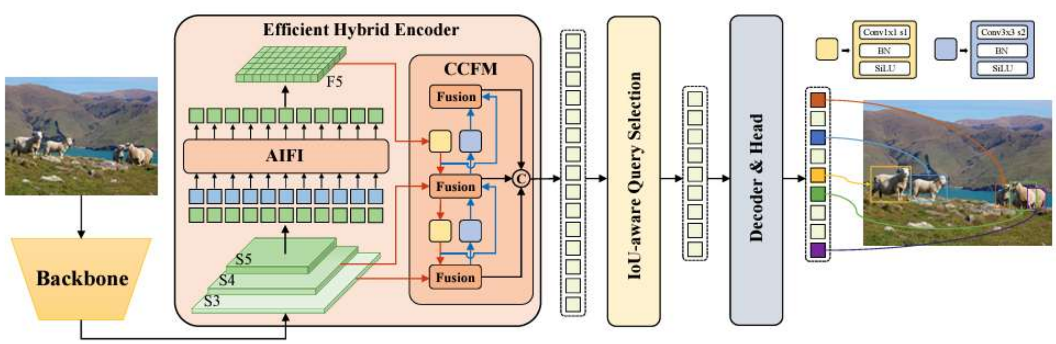
Experimental results confirm that the system achieves high-accuracy recognition and segmentation of moving objects, regardless of whether the camera is stationary or in motion. Table 1& 2 shows comparison of different object detection and tracking techniques. Fig. 5 represents architecture of dif-ferent object recognition models.

TABLE I: EVALUATION OF VARIOUS METHODS FOR RECOGNIZING OBJECTS

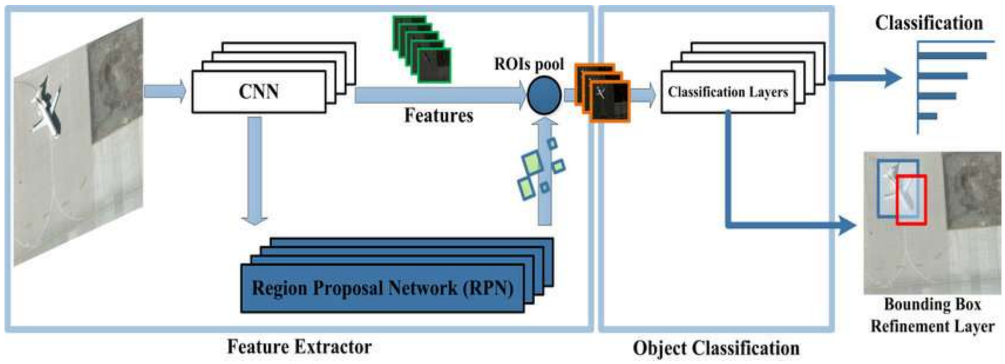
Study/ Author & Year	**Techniques**	**Features**					
		Type	**Speed**	**Accuracy**	**Strengths**	**Weaknesses**	**Best Use Case**
K. Fu et al. [14] (2020)	SSD	One-stage	Fast	Low	Balanced speed and accuracy	Lower accuracy on small objects	Mobile/embedded, quick results
A. S. Talaat et al. [29] (2023)	RetinaNet	One-stage	Moderately fast	High	Strong on rare/small classes	Slower than YOLO/SSD	Retail analytics, counting, rare items/object detection
M. H. Ashraf et al. [30] (2023)	RT-DETR	Transformer	Fast	State of the art	Real-time object detection; competitive inference speeds	Longer training times compared to some YOLO models	Flexible real-time vision tasks
K. Vijiyakumar et al. [33] (2024)	Faster R-CNN	Two-stage	Slow	Very high	Accurate localization and detection, Robust for challenging settings	Resource-intensive, slow inference	Medical/research, small objects
L. Liu et al. [35] (2023)	Mask R-CNN	Two-stage	Slow	High	Object detection + instance segmentation	slow for real-time tasks, computationally expensive	Medical Imaging, Surveillance
V. K. Deshpande et al. [37] (2023)	YOLO	One-stage	Fast	High	Real-time detection	Sometimes misses small objects	Live video, drones, autonomous vehicles, mobile



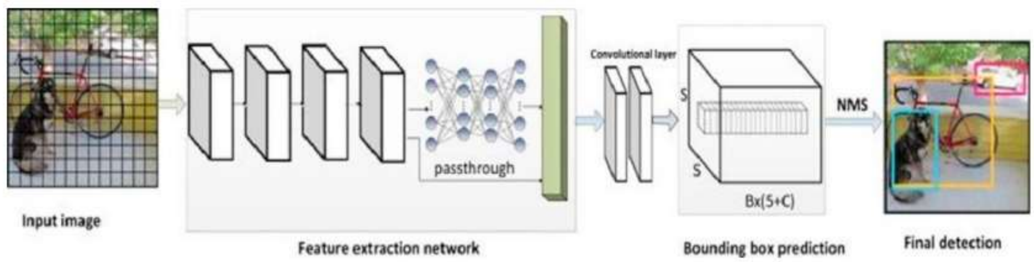
(a) RetinaNet [29]



(b) Real Time Detection Transformers (RT- DETR) [30]



(c) Faster Region-Convolutional Neural Network (R-CNN) [33]



(d) You Only Look Once (YOLO) [37]

Fig. 5: Architecture of different Object Recognition Models

TABLE II: COMPARING MULTIPLE TECHNIQUES FOR TRACKING OBJECTS

Study/Author & Year	Techniques	Features				
		Type	**Speed**	**Strengths**	**Weaknesses**	**Speed (FPS)**
K. Fu et al. [14] (2020)	SORT	Motion-based	Simple, fast, Kalman filter + Hungarian algorithm	Very fast, simple implementation	High ID switches, no appearance info	260
A. S. Talaat et al. [29] (2023)	DeepSORT	Motion + Appearance	SORT + CNN appearance features	Better ID consistency than SORT	Slower than SORT, complex setup	40-60
M. H. Ashraf et al. [30] (2023)	Byte Track	Detection-based	Associates low-confidence detections	Handles occlusions well, fast	Requires good detector	171
K. Vijiyakumar et al. [33] (2024)	Strong SORT	Enhanced DeepSORT	YOLOX detector + OSNet ReID + EMA-Kalman	State-of-the-art accuracy	Higher computational cost	30-50
L. Liu et al. [35] (2023)	YOLO Tracking	YOLO + Tracker	YOLO detection + ByteTrack/BoT-SORT	Real-time performance, easy to use	Depends on YOLO performance	30-60

IV. OBJECT DETECTION AND TRACKING DATASETS AND EVALUATION METRICS

Object recognition and tracking rely on comprehensive datasets that comprise detailed annotations. Object detection datasets include images annotated with bounding boxes, and occasionally include key points & segmentation masks [38] [39]. In contrast, object tracking datasets comprise video sequences where objects are annotated frame by frame with masks or bounding boxes.

A few popular datasets for these tasks include OpenImagesV7, ImageNet, BDD100K, KITTI, COCO (Common Objects in Context), PASCAL VOC (2007 & 2012) (Pattern Analysis, Statistics Modelling, and Computation Learning Vision Object Classes), and Multiple Object Tracking (MOT) Challenge [39]. Depending on the specific application and research goals, the dataset's choice is determined. Table 3 demonstrate comparison of existing used datasets.

TABLE III: COMPARISON OF EXISTING USED DATASETS

Study/Author	Dataset Name	Size	Classes/Categories	Annotation type	Key Feature & Use cases
A. Aslam et al. & W. Lu et al. [40] [41]	ImageNet	~14 million imag-es	Bounding boxes, la-bels	20,000 (detection: 200)	Large-scale, hierarchical classes, foundational for pre-training
A. Pandey et al. [42]	Pascal VOC	~20,000 images (VOC2007+2012)	Bounding boxes, la-bels	20	Standard for bench-marking, simple, acces-sible
T. Sharma et al. [43]	MS CO-CO	>330,000 images	Bounding boxes, masks, key points	80	Dense annotations, context-rich, segmenta-tion & captioning
M. Akteri et al. [44]	Open Images	~9 million images	Bounding boxes, masks, visual relation-ships	601	Massive scale, relation-ships, many real-world scenarios

Various key performance metrics commonly used in computer vision, especially for tasks like object detection and multiple object tracking (MOT) such as accuracy, recall, precision, maP (mean Average Precision), and several more [45]. Some of the prominent performance metric techniques are explained below:
Accuracy: The proportion of cases that were accurately predicted to all instances. It is used in Classification tasks [46].

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + Fals\ \ Negative} \quad (1)$$

Precision: How many of the predicted positives are actually correct. It is used in Classification and detection [47].

$$Precision = \frac{True\ Positive}{True\ Positive + Fal\ \ Positive} \quad (2)$$

Recall: How many actual positives are correctly identified? It is used in Classification and detection [47].

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

- True Positives (TP) = Objects of a particular class that were correctly identified
- True Negative (TN) = The quantity of cases that are truly negative and that the model accurately predicts
- False Positives (FP) = Items that are mistakenly classified as belonging to a particular class
- False Negatives (FN) = Real objects that the model missed

mean Average Precision (mAP): The mean of Average Precision (AP) across all classes. Precision and recall are taken into account by mAP for all classes and IoU criteria [48]. It is used in Object detection. There are two conditions that are

- mAP@ IoU = 0.5 means if IoU is greater than 0.5 then the identification is correct.
- The COCO standard averages mAP across several IoU brinks (0.5 to 0.95 in steps of 0.05) at mAP (mean Average Precision) @ [0.5:0.95].

$$mean\ Average\ Precision\ (mAP) = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \quad (4)$$

Where, n is the quantity of classes & AP_k is equivalent to the class k average precision.
Intersection over Union (IoU): A crucial statistic in object detection is intersection over union (IoU), which quantifies the overlap between two bounding boxes that are the ground reality bounding box and the predicted bounding box [48].

$$Intersection\ over\ Union\ (IoU) = \frac{Area\ of\ Intersection}{Area\ of\ Union} \quad (5)$$

- Intersection area: The area where the ground reality and estimate bounding boxes overlap.
- Union region: The combined region that is encompassed by the ground reality and estimated bounding boxes.

For calculating the union, this formula is used

- Union=Area of Estimated Box+Area of Ground Truth Box−Intersection Area

MOTA (Multiple Object Tracking Accuracy): Accounts for all tracking errors: FN_t (false negatives), FP_t (false positives), and IDS W_t (identity switches). It is used in multi-object tracking [49].

$$Multiple\ Object\ Tracking\ Accuracy\ (MOTA) = 1 - \frac{\sum(FN_t + FP_t + IDS\ W_t)}{\sum GT_t} \quad (6)$$

- FN_t = False Negatives
- FP_t = False Positives
- $IDS\ W_t$ = Identity Switches
- GT_t = Ground Truth objects
  - If it ranges from $-\infty$ to 1, then higher is better.

MOTP (Multiple Object Tracking Precision): Measures the precision of the estimated object locations by averaging the bounding box overlap (IoU) among ground-truth and predicted boxes. It is used in multi-object tracking [50] [51].

$$Multiple\ Object\ Tracking\ Precision\ (MOTP) = \frac{\sum_{i,t} d_{i.t}}{\sum_t c_t} \quad (7)$$

Where, $d_{i,t}$ = Distance between the object and its match (usually 1 - IoU)
c_t = Number of matches at time t

- If it is closer to 1 means more precise localization then higher is better.

V. Conclusion

In this paper, a comprehensive hybrid technique for object recognition and tracking has been studied, combining the strengths of traditional computer vision methods with the advancements of deep learning architectures. Through the integration of methods such as YOLO, Faster R-CNN, & Deep SORT, along with classic techniques like optical flow and Kalman filtering, the framework effectively addresses critical challenges including occlusion, identity switching, and environmental variability. The system works well on different sets of data and ways of testing, especially when it comes to how accurately it finds things (mAP) and how reliably it follows them over time (MOTA, MOTP).
Furthermore, the literature survey highlights the growing trend towards multi-modal and adaptive solutions that balance real-time performance with robust generalization. These methods align with this trajectory, offering a scalable and efficient solution suitable for everyday applications such as autonomous navigation, surveillance, & smart city systems. Future studies research is going to concentrate on improving computational efficiency for use on resource-constrained devices, as well as on enhancing domain adaptation and self-learning processes to increase resilience to unexpected situations.

References

[1] Y. Zhang, X. Huang, L. Wang: "Architecture design of emotion recognition early warning system based on AE-TAN algorithm", Journal of Physics: Conference Series, IOP Publishing Ltd, (2020)

[2] J. Kaur, W. Singh: "Tools, techniques, datasets and application areas for object detection in an image: a re-view, Multimedia Tools and Applications, pp. 81:38297–38351, (2022)

[3] S. Kamate, N. Yilmazer: "Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles", in Procedia Computer Science, vol. 61, pp. 436–441 (2015)

[4] S. P. Yadav, M. Jindal, P. Rani, V. H. C. de Albuquerque, C. S. Nascimento, M. Kumar: "An improved deep learning-based optimal object detection system from images", Multimedia Tools and Applications, vol. 83, pp. 30045–30072, (2024)

[5] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, W. Li: "Moving Object Detection Method via ResNet-18 with Encoder-Decoder Structure in Complex Scenes", IEEE Access, vol. 7, (2019)

[6] N. Krishnaraj, M. Elhoseny, M. Thenmozhi, M. M. Selim, K. Shankar: "Deep learning model for real-time image compression in Internet of Underwater Things (IoUT)", Journal of Real-Time Image Processing, (2019)

[7] B. Li, D. Lima: "Facial expression recognition via ResNet-50", International Journal of Cognitive Compu-ting in Engineering, vol. 2, pp. 57–64, (2021)

[8] K. Kalirajan, M. Sudha: "Moving object detection for video surveillance", Scientific World Journal, vol. 2015, pp. 907-469, (2015)

[9] Y. Tian, C. Peng, D. Wang, B. Wan: "High confidence detection for moving target in aerial video", The In-stitution of Engineering and Technology Image Processing, vol. 13, no. 14, pp. 2724–2734, (2019)

[10] Z. Li, D. H. Jiang, H. Wang, and D. Li: "Video Image Moving Target Recognition Method Based on Gener-ated Countermeasure Network", Computational Intelligence and Neuroscience, vol. 2022, pp. 7972845, (2022)

[11] Y. Wu, H. Sheng, Y. Zhang, S. Wang, Z. Xiong, W. Ke: “Hybrid Motion Model for Multiple Object Track-ing in Mobile Devices, IEEE Internet of Things Journal, vol. 10, no. 6, pp. 4735, (2023)
[12] F. Ashiq, M. Asif, M. B. Ahmad, S. Zafar, K. Masood, M. T. Mahmood, I. H. Lee: “CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People”, IEEE Access, vol. 10, (2022)
[13] Y. Sun, Z. Sun, W. Chen: “The evolution of object detection methods”, Engineering Applications of Artifi-cial Intelligence, vol. 133, pp. 108458, (2022)
[14] K. Hu, F. Lu, M. Lu, Z. Deng, Y. Liu: “A Marine Object Detection Algorithm Based on SSD and Feature Enhancement”, Complexity, vol. 2020, pp. 5476142, (2020)
[15] S. K. Pal, A. Pramanik, J. Maiti, P. Mitra, “Deep learning in multi-object detection and tracking: state of the art”, Applied Intelligence, vol. 51, no. 9, pp. 6400–6429, (2021)
[16] M. Elhoseny: “Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Vid-eo Surveillance Systems”, Circuits Systems and Signal Processing, (2019)
[17] M. Tasyurek: “SDPH: a new technique for spatial detection of path holes from huge volume high resolution raster images in near real time”, Journal of Real-Time Image Processing, pp. 21:70, (2024)
[18] M. Majcher, B. Kwolek: “Multi-Modal Pose Representations for 6-DOF Object Tracking”, Journal of Intel-ligent & Robotic Systems, pp. 110:149, (2024)
[19] S. M. M. Yousef, S. S. Mohseni, H. Dehbovid, R. Ghaderi: “A Practical Approach to Tracking Estimation Using Object Trajectory Linearization”, International Journal of Computational Intelligence Systems, pp. 17:175, (2024)
[20] M. A. Hossain, M. I. Hossain, M. D. Hossain, E. N. Huh: “DFC-D: A dynamic weight-based multiple fea-tures combination for real-time moving object detection”, Multimedia Tools and Applications, vol. 81, no. 22, pp. 32549–32580, (2022)
[21] R. B. Hatwar, S. D. Kamble, N. V Thakur, S. Kakde: “A Review on Moving Object Detection and Tracking Methods in Video”, International Journal of Pure and Applied Mathematics, vol. 118, no. 16, pp. 511–526, (2018)
[22] S. Gollapudi: “Motion Analysis and Object Tracking”, Learn Computer Vision Using OpenCV, pp. 119–145, (2019)
[23] P. Nagrath, N. Thakur, R. Jain, D. Saini, N. Sharma, J. Hemanth: “Understanding New Age of Intelligent Video Surveillance and Deeper Analysis on Deep Learning Techniques for Object Tracking”, IoT for Sus-tainable Smart Cities and Society: Internet of Things, pp. 31–63, (2022)
[24] T. Kusuma, K. Ashwini: “Real-Time Object Detection and Tracking Design Using Deep Learning with Spa-tial–Temporal Mechanism for Video Surveillance Applications”, Innovations in Computer Science and En-gineering (ICICSE 2022): Lecture Notes in Networks and Systems, vol. 565, pp. 697–705, (2023)
[25] T. Napier and I. Lee: “Using mobile-based augmented reality and object detection for real-time Abalone growth monitoring”, Computers and Electronics in Agriculture, vol. 207, pp. 107-744, (2023)
[26] M. A. Sentís, S. Velez, R. M. Pena, H. Baja, J. Valente: “Object detection and tracking in Precision Farm-ing: a systematic review”, Computers and Electronics in Agriculture, vol. 219, pp. 108-757, (2024)
[27] S. Abba, A. M. Bizi, J. A. Lee, S. Bakouri, M. L. Crespo: “Real-time object detection, tracking, and moni-toring framework for security surveillance systems”, Heliyon, vol. 10, pp. e34-922, (2024)
[28] N. H. Abdulghafoor, H. N. Abdullah: “A novel real-time multiple objects detection and tracking framework for different challenges”, Alexandria Engineering Journal, vol. 61, no. 12, pp. 9637–9647, (2022)
[29] A. S. Talaat, S. E. Sappagh: “ Enhanced aerial vehicle system techniques for detection and tracking in fog, sandstorm, and snow conditions”, The Journal of Supercomputing, Vol. 79, pp. 15868–15893, (2023)
[30] M. H. Ashraf, F. Jabeen, H. Alghamdi, M. S. Zia, M. S. Almutairi: “HVD-Net: A Hybrid Vehicle Detection Network for Vision-Based Vehicle Tracking and Speed Estimation”, Journal of King Saud University – Computer and Information Sciences, vol. 35, pp. 101657, (2023)
[31] Y. P. Huang, S. Kshetrimayum, C. T. Chiang: “Object-Based Hybrid Deep Learning Technique for Recog-nition of Sequential Actions”, IEEE Access, vol. 11, (2023)
[32] H. M. Chuang, M. Tsuchiya, S. Araki, R. Inoue, T. Ariyoshi, Y. Yasui: “Vision-based hybrid object track-ing for autonomous vehicles”, Intelligent Vehicles Symposium (IV), IEEE Access, (2024)
[33] K. Vijiyakumar, V. Govindasamy, V. Akila: “An effective object detection and tracking using automated image annotation with inception based faster R-CNN model”, International Journal of Cognitive Computing in Engineering, vol. 5, pp. 343–356, (2024)
[34] M. Salvi, U. R. Acharya, F. Molinari, K. M. Meiburger: “The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis”, Computers in Biology and Medicine vol. 128, pp. 104-129, (2021)
[35] L. Liu, B. Lin, and Y. Yang: “Moving scene object tracking method based on deep convolutional neural network”, Alexandria Engineering Journal, vol. 86, pp. 592–602, (2024)
[36] J. Ding, Z. Zhang, X. Yu, X. Zhao, Z. Yan: “A Novel Moving Object Detection Algorithm Based on Robust Image Feature Threshold Segmentation with Improved Optical Flow Estimation”, Applied Sciences (Swit-zerland), vol. 13, no. 8, (2023)
[37] S. D. Meena, V. G. S. Sameeraja, N. S. Lasya, M. Sathvika, V. Harshitha, J. Sheela: “Hybrid Neural Net-work Architecture for Multi-Label Object Recognition Using Feature Fusion”, Procedia Computer Science, vol. 215, pp. 78–90, (2022)

[38] A. Ramachandran, A. K. Sangaiah: “A review on object detection in unmanned aerial vehicle surveil-lance”, International Journal of Cognitive Computing in Engineering, vol. 2, pp. 215–228, (2021)
[39] V. K. Deshpande, R. Bhalerao, M. Chaturvedi: “Modified YOLO Module for Efficient Object Tracking”, IEEE Latin America Transactions, vol. 21, no. 3, (2023)
[40] A. Aslam, E. Curry: “A Survey on Object Detection for the Internet of Multimedia Things (IoMT) using Deep Learning and Event-based Middleware: Approaches, Challenges, and Future Directions”, Image and Vision Computing, vol. 106, pp. 104095, (2021)
[41] W. Lu, C. Lan, C. Niu, W. Liu, L. Lyu, Q. Shi, S. Wang: “A CNN-Transformer Hybrid Model Based on CSWin Transformer for UAV Image Object Detection”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, (2023)
[42] A. Pandey, R. Raja, S. Srivastava, K. Kumar, M. Gupta, C. Somthawinpongsa, A. Nanthaamornphong: “Bio-Inspired Object Detection and Tracking in Aerial Images: Harnessing Northern Goshawk Optimiza-tion”, IEEE Access, vol. 12, (2024)
[43] T. Sharma, A. Chehri, I. Fofana, S. Jadhav, S. Khare, B. Debaque, N. D. Hindie, D. Arya: “Deep Learning-Based Object Detection and Classification for Autonomous Vehicles in Different Weather Scenarios of Quebec, Canada”, IEEE Access, vol. 12, (2024)
[44] M. Akteri, M. O. Rahman, M. N. Islam, M. M. Hassan, A. Alsanad, A. K. Sangaiah: “Energy-Efficient Tracking and Localization of Objects in Wireless Sensor Networks”, IEEE Access, vol. 6, (2018)
[45] K. Tarkasis, K. Kaparis, A. C. Georgiou: “Enhancing Trustworthiness in Real Time Single Object Detec-tion”, Information Systems Frontiers, (2025)
[46] J. Xu, S. Niu, Z. Wang: “Object tracking method based on edge detection and morphology”, EURASIP Journal on Advances in Signal Processing, pp. 2024:45, (2024)
[47] D. Cao, Z. Chen, L. Gao: “An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks”, Human-centric Computing and Information Science, pp. 10:14, (2020)
[48] M. Shahid, J. J. Virtusio, Y. H. Wu, Y. Y. Chen, M. Tanveer, K. Muhammad, K. L. Hua: “Spatio-Temporal Self-Attention Network for Fire Detection and Segmentation in Video Surveillance”, IEEE Access, vol. 10, (2022)
[49] A. Khan, M. M. Fouda, D. Do, A. Almaleh, A. M. Alqahtani, A. U. Rahman: “Underwater Target Detection Using Deep Learning: Methodologies, Challenges, Applications, and Future Evolution”, IEEE Access, vol. 12, (2024)
[50] N. V. Rao, D. V. Prasad, M. Sugumaran: “Real-time video object detection and classification using hybrid texture feature extraction”, International Journal of Computers and Applications, (2018)
[51] Y. Li, H. Wang, L. M. Dang, T. N. Nguyen, D. Han, A. Lee, I. Jang, H. Moon: “Deep Learning-Based Hy-brid Framework for Object Detection and Recognition in Autonomous Driving”, IEEE Access, vol. 8, (2020)